RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _____ 10/560,224 _____

Source: _____ JFWP _____

Date Processed by STIC: _____ 12/22/2005 _____

ENTERED

CRF Errors Edited by the STIC Systems Branch

Serial Number: __10/560, 224__   CRF Edit Date: _12/22/05_

Edited by: __DA__

____ Realigned nucleic acid/amino acid numbers/text in cases where the sequence text "wrapped" to the next line

____ Corrected the SEQ ID NO. Sequence numbers edited were:

_____

____ Inserted or corrected a nucleic number at the end of a nucleic line. SEQ ID NO's edited:

_____

__✓__ Deleted: ___ invalid beginning/end-of-file text ; ___ page numbers

____ Inserted mandatory headings/numeric identifiers, specifically:

_____

____ Moved responses to same line as heading/numeric identifier, specifically:

_____

____ Other:

_____

_____

_____

Revised 09/09/2003

IFWP

```
       3 <110> APPLICANT: Lassen, Soren Flensted
       5 <120> TITLE OF INVENTION: Improved proteases and methods for producing them
       7 <130> FILE REFERENCE: 10495.204-US
C-->   9 <140> CURRENT APPLICATION NUMBER: US/10/560,224
C-->   9 <141> CURRENT FILING DATE: 2005-12-12
       9 <160> NUMBER OF SEQ ID NOS: 53
      11 <170> SOFTWARE: PatentIn version 3.3
      13 <210> SEQ ID NO: 1
      14 <211> LENGTH: 1062
      15 <212> TYPE: DNA
      16 <213> ORGANISM: Nocardiopsis sp. NRRL 18262
      19 <220> FEATURE:
      20 <221> NAME/KEY: misc_feature
      21 <222> LOCATION: (1)..(495)
      22 <223> OTHER INFORMATION: Encodes the pro-region shown in positions -165 to -1 of SEQ
ID
      23        NO:43.
      25 <220> FEATURE:
      26 <221> NAME/KEY: misc_feature
      27 <222> LOCATION: (496)..(1059)
      28 <223> OTHER INFORMATION: Encodes the mature region shown in positions 1-188 of SEQ ID
      29        NO:43.
      31 <400> SEQUENCE: 1
      32 gctactggag cattacctca gtctcctaca cctgaagcag atgcagtatc gatgcaagaa      60
      34 gcattacaac gtgatcttga tcttacatca gctgaagctg aggaattact tgctgcacaa     120
      36 gatacagcct ttgaagttga tgaagctgcc gctgaagcag ctggtgatgc atatggtggt     180
      38 tcagtattcg atactgaatc actcgaactt actgtactag tgaccgatgc agcagctgtt     240
      40 gaagctgttg aagccacagg tgcaggtaca gagctcgtat cttatggtat tgatggatta     300
      42 gatgagatcg tacaagagct taatgcagct gatgccgttc caggtgtagt tggatggtat     360
      44 cctgatgtag caggtgatac tgttgtctta gaagttcttg aaggctctgg agctgatgtt     420
      46 tctggacttt tagcagacgc aggagtcgat gcatccgcgg ttgaagtgac cacgtcagat     480
      48 cagcctgaac tctatgccga tatcattgga ggcctagcgt acacaatggg tggtcgctgc     540
      50 agcgtaggat ttgcagccac aaatgcagct ggacaacctg gcttcgtgac agctggacat     600
      52 tgcggccgcg tcggtacaca ggttactatc ggcaatggaa gaggtgtctt tgagcaaagc     660
      54 gtatttcccg ggaatgatgc tgccttcgtt agaggtacgt ccaactttac gcttactaac     720
      56 ttagtatcta gatacaacac tggcggatat gcaactgtag caggtcacaa tcaagcacct     780
      58 attggctcta gcgtctgccg ctcagggtcg actacaggat ggcattgtgg aaccattcaa     840
      60 gctagaggtc agagcgtgag ctatcctgaa ggtaccgtaa cgaacatgac tcgtacgact     900
      62 gtatgtgcag aaccaggtga ctctggaggt tcatatatca gcggtacgca agcgcaaggc     960
      64 gttacctcag gtggatccgg taactgtagg acaggtggca caacgttcta ccaggaagtg    1020
      66 acaccgatgg tgaactcttg gggagttaga ctccgtacat aa                       1062
      69 <210> SEQ ID NO: 2
      70 <211> LENGTH: 1143
      71 <212> TYPE: DNA
```

```
 72 <213> ORGANISM: Artificial sequence
 74 <220> FEATURE:
 75 <223> OTHER INFORMATION: A synthetic 10R gene (10Rsynt-15) encoding a S2A protease
denoted
 76        "10R" fused by PCR in frame to the signal peptide encoding
 77        sequence of a heterologous protease, Savinase.
 79 <400> SEQUENCE: 2
 80 atgaagaaac cgttggggaa aattgtcgca agcaccgcac tactcatttc tgttgctttt        60
 82 agttcatcga tcgcatcggc tgctactgga gcattacctc agtctcctac acctgaagca       120
 84 gatgcagtat cgatgcaaga agcattacaa cgtgatcttg atcttacatc agctgaagct       180
 86 gaggaattac ttgctgcaca agatacagcc tttgaagttg atgaagctgc cgctgaagca       240
 88 gctggtgatg catatggtgg ttcagtattc gatactgaat cactcgaact tactgtacta       300
 90 gtgaccgatg cagcagctgt tgaagctgtt gaagccacag gtgcaggtac agagctcgta       360
 92 tcttatggta ttgatggatt agatgagatc gtacaagagc ttaatgcagc tgatgccgtt       420
 94 ccaggtgtag ttggatggta tcctgatgta gcaggtgata ctgttgtctt agaagttctt       480
 96 gaaggctctg agctgatgt ttctggactt ttagcagacg caggagtcga tgcatccgcg       540
 98 gttgaagtga ccacgtcaga tcagcctgaa ctctatgccg atatcattgg aggcctagcg       600
100 tacacaatgg gtggtcgctg cagcgtagga tttgcagcca caaatgcagc tggacaacct       660
102 ggcttcgtga cagctggaca ttgcggccgc gtcggtacac aggttactat cggcaatgga       720
104 agaggtgtct ttgagcaaag cgtatttccc gggaatgatg ctgccttcgt tagaggtacg       780
106 tccaacttta cgcttactaa cttagtatct agatacaaca ctggcggata tgcaactgta       840
108 gcaggtcaca atcaagcacc tattggctct agcgtctgcc gctcagggtc gactacagga       900
110 tggcattgtg gaaccattca agctagaggt cagagcgtga gctatcctga aggtaccgta       960
112 acgaacatga ctcgtacgac tgtatgtgca gaaccaggtg actctggagg ttcatatatc      1020
114 agcggtacgc aagcgcaagg cgttacctca ggtggatccg gtaactgtag gacaggtggc      1080
116 acaacgttct accaggaagt gacaccgatg gtgaactctt ggggagttag actccgtaca      1140
118 taa                                                                    1143
121 <210> SEQ ID NO: 3
122 <211> LENGTH: 8
123 <212> TYPE: PRT
124 <213> ORGANISM: Artificial sequence
126 <220> FEATURE:
127 <223> OTHER INFORMATION: C-terminal amino acid tail expressed as fusion to protease
of the
128        invention.
130 <400> SEQUENCE: 3
132 Gln Ser His Val Gln Ser Ala Pro
133 1               5
136 <210> SEQ ID NO: 4
137 <211> LENGTH: 24
138 <212> TYPE: DNA
139 <213> ORGANISM: Artificial sequence
141 <220> FEATURE:
142 <223> OTHER INFORMATION: Polynucleotide encoding a C-terminal amino acid tail
expressed as
143        fusion to protease of the invention.
145 <400> SEQUENCE: 4
146 caatcgcatg ttcaatccgc tcca                                              24
149 <210> SEQ ID NO: 5
150 <211> LENGTH: 4
151 <212> TYPE: PRT
152 <213> ORGANISM: Artificial sequence
```

RAW SEQUENCE LISTING   DATE: 12/22/2005
PATENT APPLICATION: **US/10/560,224**   TIME: 12:16:47

Input Set : **A:\pto.da.txt**
Output Set: **N:\CRF4\12222005\J560224.raw**

```
154 <220> FEATURE:
155 <223> OTHER INFORMATION: C-terminal amino acid tail expressed as fusion to protease
of the
156          invention.
158 <400> SEQUENCE: 5
160 Gln Ser Ala Pro
161 1
164 <210> SEQ ID NO: 6
165 <211> LENGTH: 12
166 <212> TYPE: DNA
167 <213> ORGANISM: Artificial sequence
169 <220> FEATURE:
170 <223> OTHER INFORMATION: Polynucleotide encoding a C-terminal amino acid tail
expressed as
171          fusion to protease of the invention.
173 <400> SEQUENCE: 6
174 caatcggctc ct                                                      12
177 <210> SEQ ID NO: 7
178 <211> LENGTH: 2
179 <212> TYPE: PRT
180 <213> ORGANISM: Artificial sequence
182 <220> FEATURE:
183 <223> OTHER INFORMATION: C-terminal amino acid tail expressed as fusion to protease
of the
184          invention.
186 <400> SEQUENCE: 7
188 Gln Pro
189 1
192 <210> SEQ ID NO: 8
193 <211> LENGTH: 6
194 <212> TYPE: DNA
195 <213> ORGANISM: Artificial sequence
197 <220> FEATURE:
198 <223> OTHER INFORMATION: Polynucleotide encoding a C-terminal amino acid tail
expressed as
199          fusion to protease of the invention.
201 <400> SEQUENCE: 8
202 caacca                                                             6
205 <210> SEQ ID NO: 9
206 <211> LENGTH: 1
207 <212> TYPE: PRT
208 <213> ORGANISM: Artificial sequence
210 <220> FEATURE:
211 <223> OTHER INFORMATION: C-terminal amino acid tail expressed as fusion to protease
of the
212          invention.
214 <400> SEQUENCE: 9
216 Pro
217 1
220 <210> SEQ ID NO: 10
221 <211> LENGTH: 3
222 <212> TYPE: DNA
223 <213> ORGANISM: Artificial sequence
```

225 <220> FEATURE:

```
226 <223> OTHER INFORMATION: Polynucleotide encoding a C-terminal amino acid tail
expressed as
227       fusion to protease of the invention.
229 <400> SEQUENCE: 10
230 cca                                                                    3
233 <210> SEQ ID NO: 11
234 <211> LENGTH: 45
235 <212> TYPE: DNA
236 <213> ORGANISM: Artificial sequence
238 <220> FEATURE:
239 <223> OTHER INFORMATION: Primer #252639
241 <400> SEQUENCE: 11
242 catgtgcatg tgggtaccgc aacgttcgca gatgctgctg aagag          45
245 <210> SEQ ID NO: 12
246 <211> LENGTH: 44          .
247 <212> TYPE: DNA
248 <213> ORGANISM: Artificial sequence
250 <220> FEATURE:
251 <223> OTHER INFORMATION: Primer #251992
253 <400> SEQUENCE: 12
254 catgtgcatg tggtcgaccg attatggagc ggattgaaca tgcg          44
257 <210> SEQ ID NO: 13
258 <211> LENGTH: 44
259 <212> TYPE: DNA
260 <213> ORGANISM: Artificial sequence
262 <220> FEATURE:
263 <223> OTHER INFORMATION: Primer #179541
265 <400> SEQUENCE: 13
266 gcgttgagac gcgcggccgc gagcgccgtt tggctgaatg atac          44
269 <210> SEQ ID NO: 14
270 <211> LENGTH: 43
271 <212> TYPE: DNA
272 <213> ORGANISM: Artificial sequence
274 <220> FEATURE:
275 <223> OTHER INFORMATION: Primer #179542
277 <400> SEQUENCE: 14
278 gcgttgagac agctcgagca gggaaaaatg gaaccgcttt ttc          43
281 <210> SEQ ID NO: 15
282 <211> LENGTH: 64
283 <212> TYPE: DNA
284 <213> ORGANISM: Artificial sequence
286 <220> FEATURE:
287 <223> OTHER INFORMATION: Primer #179539
289 <400> SEQUENCE: 15
290 ccatttgatc agaattcact ggccgtcgtt ttacaaccat tgcggaaaat agtcataggc          60
292 atcc                                                                    64
295 <210> SEQ ID NO: 16
296 <211> LENGTH: 60
297 <212> TYPE: DNA
298 <213> ORGANISM: Artificial sequence
```

Input Set : **A:\pto.da.txt**
Output Set: **N:\CRF4\12222005\J560224.raw**

```
300 <220> FEATURE:
301 <223> OTHER INFORMATION: Primer #179540
303 <400> SEQUENCE: 16
304 ggatccagat ctggtacccg ggtctagagt cgacgcggcg gttcgcgtcc ggacagcaca      60
307 <210> SEQ ID NO: 17
308 <211> LENGTH: 37
309 <212> TYPE: DNA
310 <213> ORGANISM: Artificial sequence
312 <220> FEATURE:
313 <223> OTHER INFORMATION: Primer #179154
315 <400> SEQUENCE: 17
316 gttgtaaaac gacggccagt gaattctgat caaatgg                               37
319 <210> SEQ ID NO: 18
320 <211> LENGTH: 37
321 <212> TYPE: DNA
322 <213> ORGANISM: Artificial sequence
324 <220> FEATURE:
325 <223> OTHER INFORMATION: Primer #179153
327 <400> SEQUENCE: 18
328 ccgcgtcgac actagacacg ggtacctgat ctagatc                               37
331 <210> SEQ ID NO: 19
332 <211> LENGTH: 22
333 <212> TYPE: DNA
334 <213> ORGANISM: Artificial sequence
336 <220> FEATURE:
337 <223> OTHER INFORMATION: Primer #317
339 <400> SEQUENCE: 19
340 tggcgcaatc ggtaccatgg gg                                               22
343 <210> SEQ ID NO: 20
344 <211> LENGTH: 40
345 <212> TYPE: DNA
346 <213> ORGANISM: Artificial sequence
348 <220> FEATURE:
349 <223> OTHER INFORMATION: Primer #139 NotI
351 <400> SEQUENCE: 20
352 catgtgcatg cggccgcatt aacgcgttgc cgcttctgcg                            40
355 <210> SEQ ID NO: 21
356 <211> LENGTH: 7443
357 <212> TYPE: DNA
358 <213> ORGANISM: Artificial sequence
360 <220> FEATURE:
361 <223> OTHER INFORMATION: Sequence of plasmid pMB1508
363 <400> SEQUENCE: 21
364 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca      60
366 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg     120
368 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc     180
370 accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc     240
372 attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat     300
374 tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt     360
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 12/22/2005
PATENT APPLICATION: **US/10/560,224**   TIME: 12:16:48

Input Set : **A:\pto.da.txt**
Output Set: **N:\CRF4\12222005\J560224.raw**

<u>**Please Note:**</u>
**Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.**

Seq#:24; N Pos. 13,16

VERIFICATION SUMMARY   DATE: 12/22/2005
PATENT APPLICATION: **US/10/560,224**   TIME: 12:16:48

Input Set : **A:\pto.da.txt**
Output Set: **N:\CRF4\12222005\J560224.raw**

L:9 M:270 C: Current Application Number differs, Replaced Current Application No
L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:849 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:24 after pos.:0

Raw Sequence Listing before editing,
for reference only

IFW ρ

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/560,224

DATE: 12/20/2005
TIME: 10:21:35

Input Set : **A:\01-SQ Listing-12 Dec 2005.txt**
Output Set: **N:\CRF4\12202005\J560224.raw**

```
    3 <110> APPLICANT: Lassen, Soren Flensted
    5 <120> TITLE OF INVENTION: Improved proteases and methods for producing them
    7 <130> FILE REFERENCE: 10495.204-US
C-->  9 <140> CURRENT APPLICATION NUMBER: US/10/560,224
C-->  9 <141> CURRENT FILING DATE: 2005-12-12
    9 <160> NUMBER OF SEQ ID NOS: 53
   11 <170> SOFTWARE: PatentIn version 3.3
```

Does Not Comply
Corrected Diskette Needed

$(\text{P9}-2)$

ERRORED SEQUENCES

```
 2425 <210> SEQ ID NO: 53
 2426 <211> LENGTH: 166
 2427 <212> TYPE: PRT
 2428 <213> ORGANISM: Artificial sequence
 2430 <220> FEATURE:
 2431 <223> OTHER INFORMATION: Shuffled propeptide G-1.2
 2434 <220> FEATURE:
 2435 <221> NAME/KEY: PROPEP
 2436 <222> LOCATION: (1)..(166)
 2438 <400> SEQUENCE: 53
 2440 Ala Thr Gly Ala Leu Pro Gln Ser Pro Thr Pro Glu Ala Asp Ala Val
 2441 1               5                   10                  15
 2444 Ser Met Gln Glu Ala Leu Gln Arg Asp Leu Asp Leu Thr Ser Ala Glu
 2445              20                  25                  30
 2448 Ala Glu Glu Leu Leu Ala Ala Gln Asp Thr Ala Phe Glu Val Asp Glu
 2449          35                  40                  45
 2452 Ala Ala Ala Ala Ala Ala Gly Asp Ala Tyr Gly Gly Ser Ile Phe Asp
 2453      50                  55                  60
 2456 Thr Glu Thr Leu Glu Leu Thr Val Leu Val Thr Asp Ser Ser Ser Val
 2457 65                  70                  75                  80
 2460 Glu Ala Val Glu Ala Ala Gly Ala Glu Ala Lys Val Val Ser His Gly
 2461              85                  90                  95
 2464 Met Glu Gly Leu Glu Glu Ile Val Ala Asp Leu Asn Ala Ala Asp Ala
 2465             100                 105                 110
 2468 Gln Pro Gly Val Val Gly Trp Tyr Pro Asp Ile His Ser Asp Thr Val
 2469             115                 120                 125
 2472 Val Leu Glu Val Leu Glu Gly Ser Gly Ala Asp Val Asp Ser Leu Leu
 2473         130                 135                 140
 2476 Ala Gly Ala Gly Val Asp Thr Ala Asp Val Lys Val Glu Ser Thr Thr
 2477 145                 150                 155                 160
 2480 Glu Gln Pro Glu Leu Tyr
 2481                 165
```

RAW SEQUENCE LISTING   DATE: 12/20/2005
PATENT APPLICATION: **US/10/560,224**   TIME: 10:21:35

Input Set : **A:\01-SQ Listing-12 Dec 2005.txt**
Output Set: **N:\CRF4\12202005\J560224.raw**

E--> 2487/44

deleted

VERIFICATION SUMMARY   DATE: 12/20/2005
PATENT APPLICATION: **US/10/560,224**   TIME: 10:21:36

Input Set : **A:\01-SQ Listing-12 Dec 2005.txt**
Output Set: **N:\CRF4\12202005\J560224.raw**

L:9 M:270 C: Current Application Number differs, Replaced Current Application No
L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:849 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:24 after pos.:0
L:2487 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:53